Representing Protein and Peptide Structures with Parallel-Coordinates

OREN M. BECKER

Department of Chemical Physics, School of Chemistry, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

Received 4 April 1997; accepted 20 June 1997

ABSTRACT: Graphical representation of molecular conformations is an important tool used by chemists to gain molecular insight. In spite of today's enhanced computer graphics there are still situations, such as in multiple conformation displays, in which standard visualization techniques are limited. Parallel-coordinate (||-coords) representation, which was originally developed for visualizing multivariant datasets in fields other than chemistry, offers an alternative basis for graphical representation of molecular structures. In parallel-coordinates, the axes are drawn parallel rather than perpendicular to each other, allowing many axes to be placed and seen. This mapping procedure has unique geometric properties and useful relationships to the original space. In this article, we apply the parallel-coordinate representation for presenting peptide and protein structural conformations. In particular, we demonstrate the usefulness of parallel-coordinates in the context of conformational analysis where this representation, combined with multiple filters, allows nontrivial clustering of data points, leading to new observations. The ||-coords representation is also demonstrated as a tool for two-dimensional (2D) representation of protein secondary structure and for identification of disulfide-bonded pairs in protein structures. Regardless of the application, an advantage of the ||-coords approach is that it retains its inherent simplicity and ease of use, and requires little or no software development. *J Comput Chem* **18**: 1893–1902, 1997

Keywords: parallel coordinates; visualization; peptides; proteins; conformational analysis; secondary structure; disulfide bonds

Correspondence to: O. M. Becker; e-mail: becker@sapphire.tau.ac.il; http://www.tau.ac.il/~becker

Contract/grant sponsor: Tel Aviv University

CCC 0192-8651 / 97 / 151893-10

Introduction

Graphical representation of molecular conformations has accompanied chemistry from its earliest days. Whether it is an inorganic complex, a complicated organic molecule, or a large protein chemists have always resorted to visual aids to gain insight into molecular structure and properties. The availability of computer graphics has further enhanced the visualization capabilities allowing unprecedented possibilities. Nevertheless, even enhanced computer graphics has limitations. Such is the case, for example, when multiple conformations are involved. Overlaying many molecular conformations one on top of the other in a single display (e.g., multiple NMR structures) can at most yield only a qualitative notion of the extent of their similarity, provided that the structures do not differ too much. If the structures are significantly different the visual result may be unintelligible and it is necessary to resort to tools, such as cluster analysis, which selects a small subset of conformations for further study.[1]

Such a situation often arises in the context of conformational analysis, which is a basic computational procedure used in a variety of biophysical studies. It is used, for example, to search for the most stable structure of small- to medium-sized molecules,[2,3] to analyze molecular flexibility (e.g., ref. 1), and in the context of designing new drugs.[4] The basic idea of conformational analysis is to analyze large numbers of molecular conformations to identify correlations between molecular structure and molecular properties, such as the potential energy. Several nonvisual analysis procedures are available for conformational analysis (see ''Conformation Analysis'' section), but this field is a clear example where *visual analysis* of the whole set is practically useless. Overlaying hundreds of conformations in a single display will typically result in a complete loss of information.

The main problem with displaying many conformations of the same N-atom molecule in one picture stems from the fact that standard three-dimensional Cartesian coordinates are usually used to represent the molecule. This means that each conformation adds N objects (atoms) to the display, resulting in a very large number of objects when simultaneously displaying the whole conformation ensemble. The large number of objects, even when using a subset selected by cluster analysis, often makes it very hard to extract information based on visual inspection of the plot.

In this study, a different approach, based on the parallel-coordinates representation,[5–7] is used to obtain graphical representations of molecular structures. Following a brief discussion of the parallel-coordinate representation (next section), we demonstrate its application to polypeptide conformations (third section), to peptide conformational analysis (fourth section), to the graphical representations of protein secondary structure (fifth section), and for visual identification of disulfide-bonded pairs in proteins (final section).

Parallel Coordinates

''Parallel coordinates'' (||-coords) is a method designed for visualizing complex multivariant datasets and multidimensional systems.[5–7] In this representation, the axes are drawn *parallel* rather than perpendicular to each other as shown in Fig. 1, allowing many axes to be placed and to be seen. Parallel coordinates induce a 1:1 mapping between subsets of N-space into two dimensions (2D), where each point in N-space is represented by a line in 2D. This mapping is not a projection (such as in principal coordinate analysis[8,9]) and does not loose information in the process. What distinguishes ||-cords from other approaches is the ability to represent and display relations between the

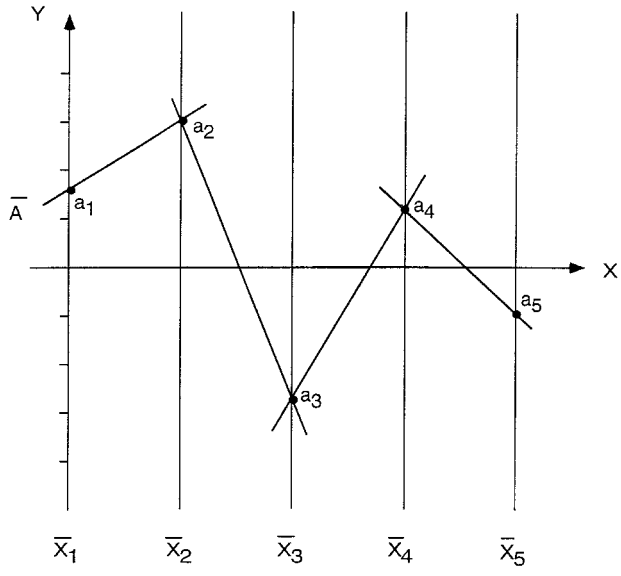

FIGURE 1. The polygonal line $\bar{A}$ is the parallal-coordinate representation of the point $A = (a_1, a_2, a_3, a_4, a_5)$.

different variables. These are typically represented as the envelope of the family of lines that correspond to the points of the relation.

An example that demonstrates some of the properties of the ||-coords representation is the point $\leftrightarrow$ line duality in 2D ||-coords.[6] A point in the $\{X_1 X_2\}$-plane is represented by a segment between the $\overline{X}_1$ and $\overline{X}_2$ axes in parallel coordinates (the distance between the parallel axes in the embedding xy-plane is d) and, in fact, by the *line* containing the segment. The line, l: $X_2 = mX_1 + b$, is an infinite collection of points A. These points are represented by the infinite collection of lines $\overline{A}$ on the $\{\overline{X}_2\ \overline{X}_2\}$ ||-coords plane, which, when $m \neq 1$, intersect at a *point* $\bar{l}$ with xy coordinates $[d/(1-m), b/(1-m)]$. This means that, for $m < 0$, the line segments will cross between $\overline{X}_1$ and $\overline{X}_2$ (Fig. 2a), whereas for $m > 0$ (and $m \neq 1$) the line segments meet at a point outside these two parallel axes (Fig. 2b). For $0 < m < 1$, the point $\bar{l}$ is to the right of $\overline{X}_2$ and for $m > 1$ it is to the left of $\overline{X}_1$. The point $\bar{l}$, which is the *envelope* of the family of lines $\overline{A}$, suffices to represent, all by itself, the linear relation of l, as the two parameters m and b specify l completely. If $m = 1$ the line segments in ||-coords will be parallel. On the other hand, a family of parallel lines in Cartesian coordinates is represented in ||-coords by a family of points on a vertical line.

The role of envelopes is even more evident in the ||-coords representation of a curve, r. Such a curve is represented by the set of lines (an envelope) which corresponds to the points of r. An example of such a relationship is the mapping of a two-dimensional hyperbolic curve into an ellipse envelope in ||-coords, which is a special case of a general result involving convex sets.[5] These dualities generalize nicely to ||-coords representations of N-dimensional spaces; for instance, a line in N-dimensions can be uniquely represented by $N-1$ points in ||-coords. The ||-coords transformation has additional geometric properties that allow representations of objects such as hypersurfaces, identify colinearity and coplanarity, and find points within convex hypersurfaces.[5,7]

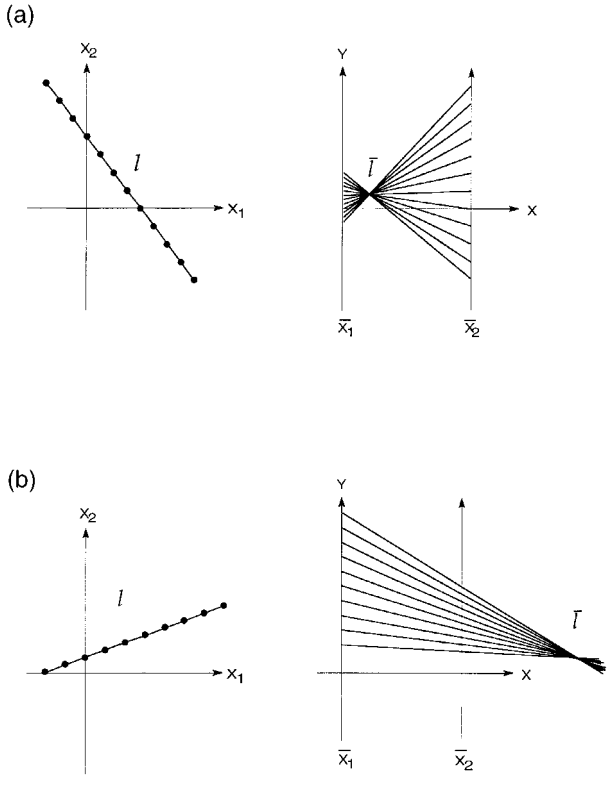

FIGURE 2. The point $\leftrightarrow$ line duality in two-dimensional ||-coords. (a) A line with a slope $m < 0$; (b) a line with a slope $0 < m < 1$.

||-Coords Representation of Molecular Conformations

So far, the most widespread application of ||-coords is ''visual data mining'' of multivariant datasets in fields other than chemistry.[10] Previously, only one study was reported on using ||-coords for representing molecular conformations.[11] In this study, we show how ||-coords can be used to represent a variety of molecular properties in polypetides, ranging from conformation clustering to identification of secondary structure and disulfide-bonded pairs.

The first, and most straightforward, application of ||-coords to molecular systems is representation of individual molecular conformations. As discussed previously, parallel coordinates map a point in $N-$dimensional space to a polygonal line in 2D. That is, polypeptide conformations defined as points in the $3N-6$ conformation space can be represented as a $3N-7$ segmented polygon in 2D. This, however, may be a very long polygon as N is often ≥ 100. A more useful plot may be obtained by applying the ||-coords representation to *subsets* of the full coordinate space. This subset can be defined either in terms of Cartesian coordinates (such as backbone C_α coordinates) or in terms of internal coordinates (e.g., dihedral angles). The different representations may be useful in different contexts.

Conformations of polypetide chains are often described by their main-chain dihedral angles (ϕ, ψ). These dihedral-angle pairs, defined by the atom quartets N—C_α—C—N and C—N—C_α—C, give the local backbone conformation of each participating amino acid around its C_α atom. As an example, Figure 3 shows the ||-coords, representation of the lowest energy conformation of the tetrapeptide isobutyryl-$(ala)_3$-NH-methyl (IAN) based on its seven main-chain dihedral angles (Fig. 4).

A similar approach was used by Luke[11] who generated ||-coords representations for the conformations of the pentapeptide Met-enkephalin. In that work ||-coords plots, drawn in terms of *all* the dihedral angles (backbone and side-chain dihedral angles), were used for analyzing a geometry minimization procedure. While informative in that context, the all-dihedral ||-coords representation may have drawbacks in other contexts. For example, using all the dihedrals is less useful for exploring the overall molecular structure, because the backbone conformation is masked by the side-chain dihedral angles. In addition, the all-dihedral ||-coords plot becomes rapidly crowded when applied to large molecules such as proteins.

A potential drawback of any ||-coords representation based on dihedral angles is the problem associated with mapping a periodic function on a one-dimensional line. This causes the top and bottom of each parallel axis to be identical resulting in a ''wraparound'' effect, in which a small change may move a data point from the top of the axis to its bottom. To overcome this problem we choose an axis-range assignment that minimizes the unwanted effects (discussed later). Another option, suggested by Luke,[11] is to wrap the parallel axes on a cone, resulting in a concentric coordinate plot. Each coordinate is represented by a circle of a different radius. These concentric representations are suitable when a small number of dihedrals is involved (around 10 angles). For larger systems, the circles become too large or their spacing becomes too narrow.

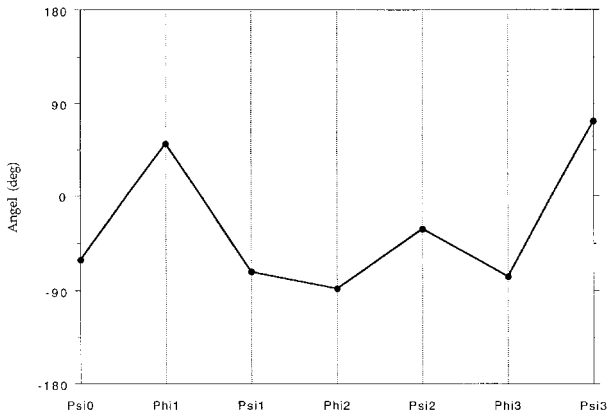

FIGURE 3. The ||-coords representation of the lowest energy conformation of the tetrapeptide isobutyrl-$(ala)_3$-NH-methyl (IAN), based on its seven main-chain ϕ, ψ diehdral angles.

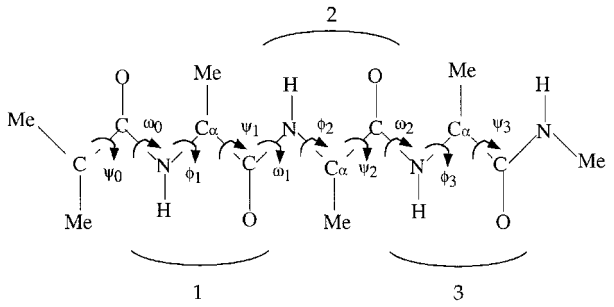

FIGURE 4. The tetrapeptide IAN [isobutyryl-$(ala)_3$-NH-methyl]. The soft torsions (ϕ, ψ) are on each side of the $C(i)$ carbons. The ω torsion angle is associated with the peptide bond.

Finally, it should be noted that, although the parallel-coordinate plots stems from a rigorous mathematical background, it conforms to a long tradition of representing polypetide properties as a function of their sequence. Examples of these include the hydropathy plots for membrane proteins,[12] rms deviations during protein dynamics,[13] and many others. Of course these plots lack the geometrical interpretation of the ||-coords plots and do not have the option of adding property-axes to the sequence axes (see next section).

Conformation Analysis

Analyzing molecular conformation spaces, even for a relatively small molecule, is a very demanding computational task, because such spaces are of extremely large dimensions. A molecule of N atoms has $3N$ degrees of freedom, and its conformation space is $3N - 6$ dimensional. As a result, even small peptides have very large conformation spaces (e.g., seven residue peptides have conformational spaces of about 100–150 dimensions). Understanding properties in such high-dimensional spaces, which is the task of conformational analysis, is therefore very complicated. In view of the importance of conformational analysis, as a basic procedure used in many computational bio-

physical studies, it is surprising that the number of computational tools available for performing such an analysis is very limited.

Conformational analysis can be performed on large samples of molecular conformations ''collected'' in any number of ways. Two common sampling procedures are quenched high-temperature molecular dynamics trajectories[1, 14, 15] and Monte Carlo sampling followed by minimization.[1, 16, 17] The resulting sets of conformations are usually sorted according to potential energies and ordered based on pairwise distances between them (e.g., the root-mean-square distance between two conformations). Interconformation distances are compiled into a ''distance matrix,'' Δ, where the elements Δ_{ij} are the rms distances between conformation i and conformation j. The distance matrix can be further analyzed by clustering algorithms that group together similar conformations, from which individual representative structures are selected for further analysis (e.g., ref. 1). Recently, more detailed conformation analysis approaches have been proposed, including ''topological mapping'' by Becker and Karplus,[18] statistical analysis by Kunz and Barry,[19] 3D projected surfaces by Becker,[20] and various applications of principal coordinate analysis.[8, 9, 21–23]

Parallel coordinate representation constitutes a new and powerful conformation analysis tool. Several properties make the ||-coords representation especially suitable for conformational analysis. The first is the ease of representing multiple conformations on the same plot. Each conformation is mapped onto a simple polygonal line in 2D so that drawing many such lines in a single plot is very easy. Moreover, the resulting plot retains the simplicity and clarity of the single conformation plot, distinguishing it from standard Cartesian plots which become unintelligible if too many conformations are overlaid. Second, unlike standard geometrical representations, ||-coords allow *mixing* geometrical coordinate axes with molecular property axes (e.g., potential energy) on the same plot. This combination offers an easy way for identifying relationships between coordinates and molecular properties. A third advantage is the possibility of making dynamic clustering and filtering choices, based on visual analysis of the patterns in the plot. These may be quite hard to reproduce by standard numerical alogrithms that apply strict cutoff criteria.

Figure 5a shows *all* 139 conformations of IAN tetrapeptide[18, 24] using an 8-axis ||-coords representation. Of these axes, seven correspond to ϕ, ψ backbone dihedral angles (similar to Fig. 3) whereas the eighth is the potential energy of that conformation. Visual inspection of the resulting picture allows one to draw conclusions that would otherwise require tedious numerical analysis. For example, clustering patterns are immediately visible in this plot. By inspecting the ''Phi3'' axis we see that all of the conformations fall into one of four possible ϕ_3 values (in the vicinity of $-160°$, $-80°$, $60°$, and $170°$; the last one is probably a wraparound of the first) and that each value cluster has a different width. Very similar clustering patterns are seen on the other two ''Phi'' axes. Clustering is less evident on the ''Psi'' axes, with the exception of ''Psi0,'' the N-terminal dihedral, which shows a two-value distribution: $-60°$ and $+120°$, related by a $180°$ rotation. These observations are similar to previous observations derived from detailed numerical analysis of this set of conformations.[18] Considering the energy axis, we see that the conformations are spread over a 12-kcal/mol range (energy is measured relative to the lowest energy conformation) and that there are no clear clustering pattern on the energy axis.

To gain more insight one can use the ||-coords plot to focus on specific regions of interest, and thus identify otherwise obscure correlations. Such focusing, or *filtering*, can be obtained by selecting only those lines (i.e., conformations) that pass through a given ''value window'' on any of the axes, whether a geometry axis or a property axis. Finer and more restrictive focusing can be obtained by applying multiple filters; that is, by selecting only those lines that pass through two or more ''value windows'' on different axes of the plot. As the filtering process can be done interactively it is naturally incorporated in the visual analysis approach and does not require additional tools.

Figure 5b is the ||-coords plot of Figure 5a after applying a ''Phi3'' filter. The specific filter applied in this figure selected only those lines (i.e., conformations) that share a ''Phi3'' dihedral angle in the range $+60° \pm 10°$. For this peptide system, the filter selected 45 of a total of 139 conformations. The selected conformations were clustered very tightly on the ''Phi3'' axis, with an average ϕ_3 dihedral value of $59.7°$ and a standard deviation of $2.3°$. In general, however, filtering does not have to be so tight. The resulting plot shows that none of the very low energy conformations of IAN passed through this filter, because the lowest energy through this filter is about 2 kcal/mol higher than the global minimum (1.702 kcal/mol). This confor-

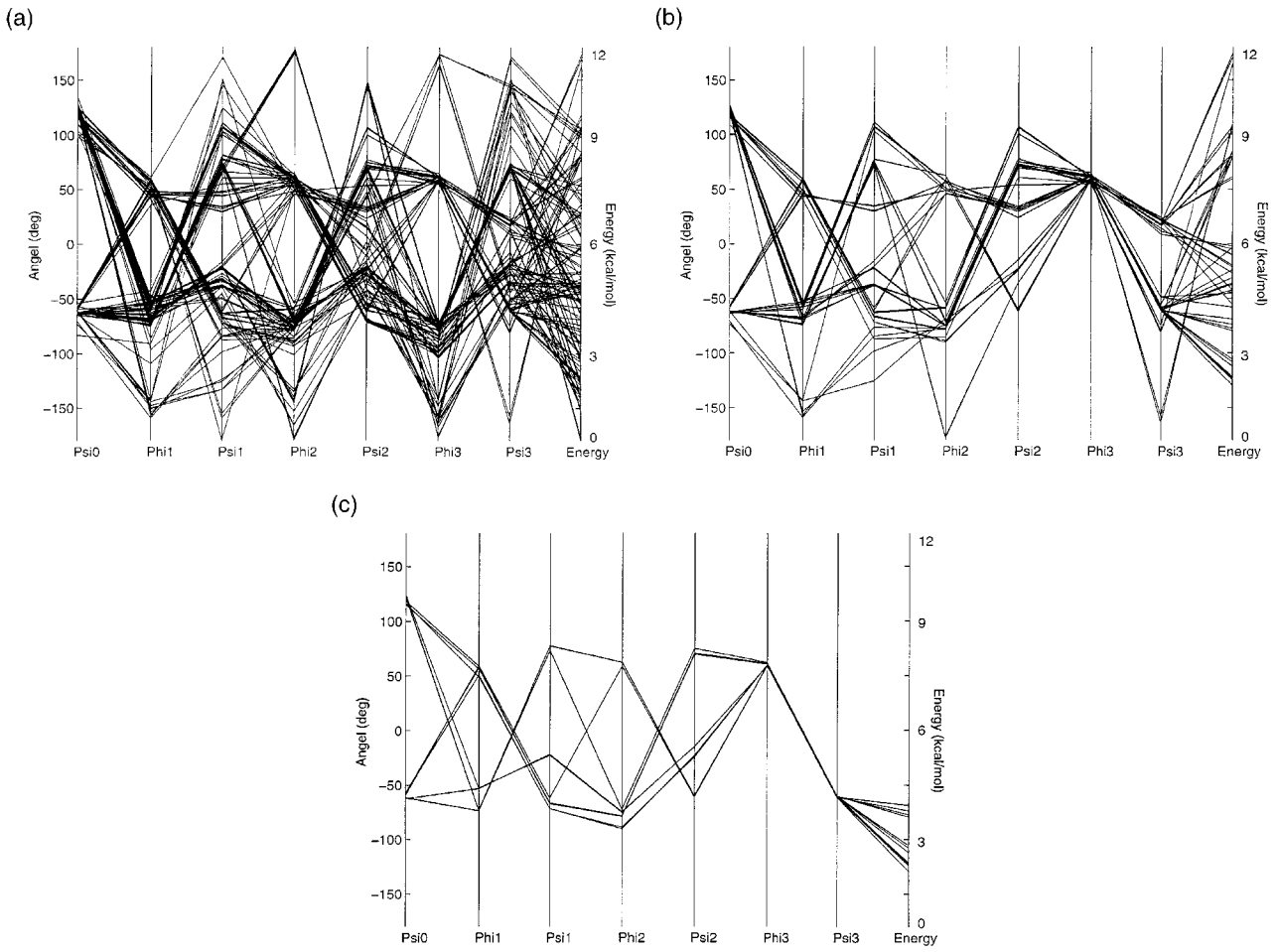

FIGURE 5. (a) An eight-axis ||-coords representation of *all* 139 conformations of the IAN tetrapeptide. Seven axes corresponds to the seven ϕ, ψ backbone dihedral angles (similar to Fig. 3) and the eighth axis, on the right-hand side of the plot, is the potential energy of the conformations. (b) Same ||-coords plot as (a) after applying a ''Phi3'' filter. The filter selects only those lines (i.e., conformations) that have a ϕ_3 value in the range $+60° \pm 10°$. (c) The same ||-coords plot as (a) after applying double filtering. This double filter has selected 12 lines (= conformations) that pass through both the $\phi_3 = 60° \pm 10°$ dihedral filter and an $E <$ 4-kcal / mol energy filter. Correlations between (ϕ, ψ) pairs are evident (see text).

mation was ranked as number 14 in the list of minima sorted by energy. Comparing Figure 5b to Figure 5a one also sees that some dihedral angle values are excluded once the $\phi_3 = 60° \pm 10°$ requirement was imposed, indicating nontrivial correlations between the different variables. For example, after applying this ϕ_3 selection, all ψ_3 values higher than 25° were also filtered out. Also affected were the ϕ_2 values, where angles higher than 65° and lower than $-90°$ were not observed (there were two exceptions with dihedral values of $-178.3°$ and $-176.9°$). It was previously noted that, for this tetrapeptide, a $\phi_3 = 60°$ value is characteristic of regions on the potential energy surface *outside* the main energy ''funnel.''[18]

The result of a *double filter* is shown in Figure 5c. In this figure, the conformations were selected by applying a second filter in addition to the one already used in Figure 5b. This double filter selects lines that pass through both the $\phi_3 = 60° \pm 10°$ dihedral filter and an $E <$ 4-kcal/mol energy filter. Only 12 of 139 conformations pass this double criteria. The first observation is that *all* 12 conformations had the same ψ_3 dihedral angle (mean $-60.2°$, standard deviation 0.6°) in addition to the same ϕ_3 value they were filtered for (mean 59.9°,

standard deviation 1.1°). We also see that the majority of the dihedral pairs ϕ_1 and ψ_1 and ϕ_2 and ψ_2 accepted the values $(-75, +75)$ or $(+60, -60)$ —values not far from those characteristic of type V or type V′ turns.

Protein Secondary Structure

Secondary structure classifications are commonly used for characterizing protein three-dimensional structures. It is thus surprising that 2D representation of these polypeptide backbone structures are rather limited. Visualizing secondary structures on computer screens, which enables 3D visualization through image rotation, is typically done by the standard cylinder (= α-helix) and arrow (= β-strand) representation or by 3D ribbons. When these images are projected into two dimensions they no longer illustrate the exact secondary structure assignments. Although they usually still convey the overall structure of the protein (e.g., a barrel) some parts of the structure are naturally obscured from sight. Graphical 2D representations that really convey the secondary structure assignments are typically based on complicated graph schemes, such as those used in PROMOTIF,[25] which shows the hydrogen-bond connectivity in protein structures. These graphical representations are highly nontrivial and require sophisticated software to generate. The need for a simple and quick way of generating 2D representations of protein secondary structure has recently led to two new suggestions. The first is a simple color-coded representation, introduced by Berendsen and collaborators[26] based on the DSSP classification,[27] which enables a time-dependent representation of protein secondary structure. The second is a mapping procedure, suggested by Smith et al.,[28] which maps the protein conformation onto a plane defined by pseudo-dihedral angles and scalar products of peptide-group vectors.

||-Coords offer a different approach for representing the protein secondary structure in 2D which is straightforward, simple, and easy to construct. This ||-coords representation is based on the fact that secondary structure elements are characterized by continuous regions of well-defined backbone (ϕ, ψ) dihedral angles, usually represented by a Ramachandran plot. Detailed analyses of model systems[29] and many actual protein structures[30] have shown that right-hand α-helices are in the third quadrant of the Ramachandran plot with values around $(-60, -60)$, β-strands cover a broad area in the second quadrant in the vicinity of $(-120, +120)$, and left-handed helics are in the first quandrant in the vicinity of $(+60, +60)$. These characteristics hold for all residues except Gly and Pro, which show different distributions of (ϕ, ψ) angles.

A ||-coords, representation, similar to the one just discussed for representing conformations of small peptides, can be applied to protein secondary structure. As before, we choose to focus on the protein's ϕ, ψ backbone dihedrals which are the most significant structure-defining coordinates. Therefore, a ||-coords plot of a conformation of an M-residue protein will consist of $2M$ parallel coordinates ordered in (ϕ, ψ) pairs along the protein's sequence. The protein conformation is represented as a polygonal line in 2D and its secondary structure elements (α-helices, β-strands, and β-turns), characterized by stretches of well-defined dihedral values, will appear as continuous stretches of similar dihedral values. Based on the periodic character of the dihedral angles, the y-axis of the ||-coords plot can be assigned any continuous range of 360°. It does not have to be assigned the standard $\{-180:+180\}$ range (previously used in Fig. 3 and 5). Because β-strands are located in the second quadrant of the Ramachandran plot (negative ϕ and positive ψ) we find it better to set the y-axis range to $\{0:+180, -180:0\}$, which is identical to $\{0:360\}$ (this choice also solves most cases of axis wraparound, discussed previously). With this axis assignment, α-helices will appear as stretches of similar angles in the upper quarter of the plot (between the values −40° and −65°) and β-strands will occupy a broad region in the midsection of the plot (between −90 and +120). Turn and coil regions, which are characterized by (ϕ, ψ) values other than those just indicated, will appear in the ||-coords plot as ''spikes,'' that is, short regions of widely varying values.

Figure 6 is a ||-coords plot of the crystal structure of apolipoprotein E4 (pdb code 1le4[31]—a predominantly α-helical protein. This protein includes five helices, at residues 24–42 (length 19), 44–53 (length 10), 54–82 (length 29), 87–122 (length 36), and 130–162 (length 33), separated by short segments representing turns. The α-helical range, between −40 and −65, is shaded. The five helices, as well as the nonhelical segments (downwards ''spikes''), are clearly seen in the ||-coords plot.

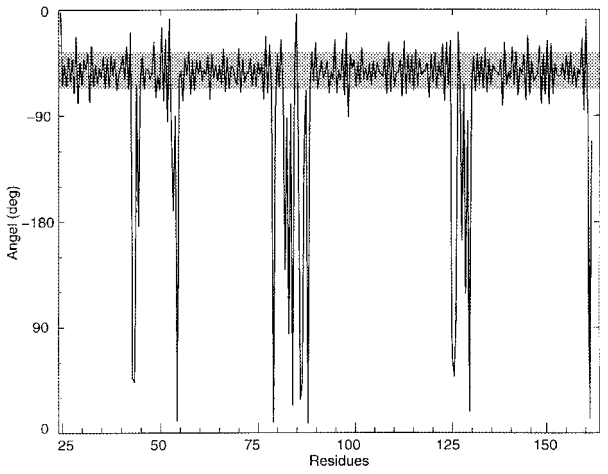

FIGURE 6. A ||-coords plot of the crystal structure of apolipoprotein E4 (1le4), a predominantly α-helical protein containing five helices. The α-helical region of the ||-coords plot (−40 to −65) is shaded. The five helices, as well as the nonhelical segments (downwards ''spikes''), are clearly seen.

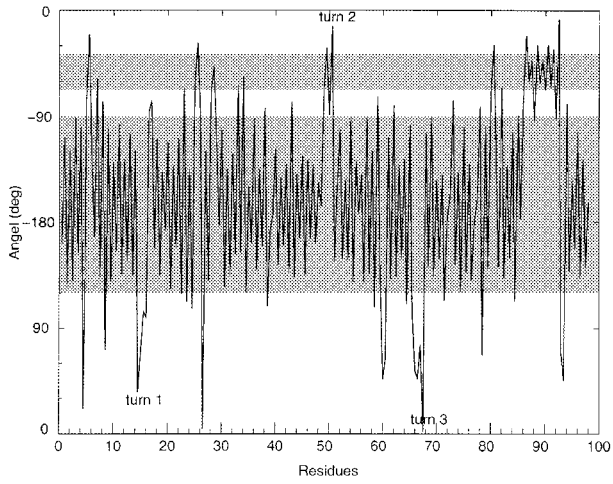

FIGURE 7. A ||-coords plot of the crystal structure of HIV-1 protease (4hvp), a predominantly β-sheet protein containing 10 β-strands and a single short α-helix. The α-helical region (−40 to −65) and the β-strand region (−90 to 120) are shaded. The three well-defined turns in this protein are also marked.

Based on this plot one sees that the third helical region is slightly shorter than reported (it ends at residue 79 rather than 82).

Figure 7 is the ||-coords plot of the crystal structure of HIV-1 protease (pdb code 4hvp[32])—a predominantly β-sheet protein. This protein (which is part of a dimer) has 10 β-strands (residues 1–4, 9–15, 18–25, 31–34, 43–49, 52–60, 65–66, 69–78, 83–85, and 96–99) and a single, short α-helix at residues 86–94 (length 9). The α-helical region (−40 to −65) and the β-strand region (−90 to 120) are shaded. The different secondary structural elements are clearly seen in this plot. All breaks between secondary structure elements appear in this plot as ''spikes,'' whereas the three well-defined β-turns (at residues 16–18, 50–52, and 66–68) stand out in the ||-coords plot as two- or three-residue structural features. As before, close inspection of the plot yields slight modifications to the assigned secondary structure ranges.

Disulfide Bonds

The ||-coordinate representation is, of course, not limited to internal coordinates (such as the dihedral coordinates used so far). It can just as well be applied directly to the Cartesian coordinates of the individual atoms. A simple example where such a representation is useful is in the context of disulfide bonds, a nonlocal structural element that has an important role in determining the protein's 3D fold. Usually, identifying bonded cysteine pairs in a new 3D protein structure requires either specific numerical measurements or interactive rotation of the 3D image on a graphics computer screen. As before, ||-coords offers a simple straightfoward visual identification of the paired S atoms.

Mapping each of the Cartesian coordinates of all protein cysteine sulfur atoms (three coordinates per atom) onto the ||-coords plane results in each atom being represented by two-segment polygons. Because the two sulfur atoms that are chemically bound must be close to each other in Cartesian space, the corresponding polygonal lines must also lay close to each other in the ||-coords plane. This means that once the ||-coords representation is drawn, it becomes a trivial visual task to identify bonded pairs; that is, pairs of lines that are close to each other in the plane. Figure 8 shows the ||-coords representation of the x, y, z coordinates of the *six* cysteine sulfur atoms in the protein BPTI (bovine pancreas trypsin inhibitor) based on the crystal structure of Marquart et al. (residues 5 and 55, 14 and 38, 30 and 51).[33] The six coordinate sets were taken from the crystal structure files as is, without any manipulation. The three S—S bonded pairs are immediately evident. Moreover, if, as sometimes happens, some cysteine S atoms are not

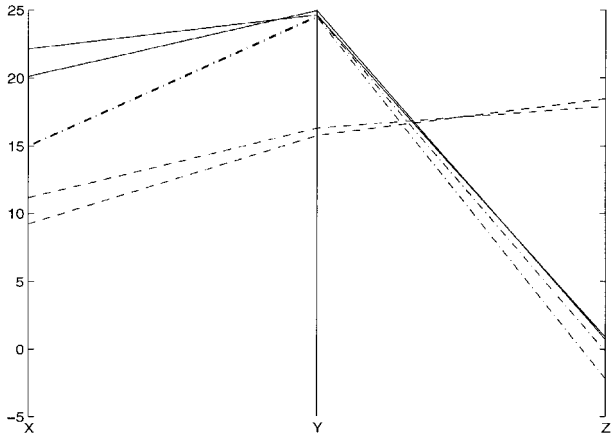

FIGURE 8. The ||-coords representation of the 3D coordinates the six cysteine sulfur atoms in the crystal structure of the protein BPTI. The three atom pairs involved in disulfide bonds are evident.

involved in disulfide bonding this will be immediately evident in the ||-coords representation, as these atoms will show up as unpaired polygons.

Summary

We have shown that the general mapping approach of parallel-coordinates (||-coords), which was previously applied mainly for visual data mining in fields other than chemistry, has many natural applications in chemistry. In particular, this was demonstrated with regard to polypeptide and protein structural analysis. Whether it is used to represent a single molecular conformation (e.g., ref. 11), to carry out conformational analysis, or to enable graphical representation of protein secondary structure, the ||-coords approach retains its natural simplicity and ease of use. Moreover, in all these cases, using ||-coords representation required few or no software developments.

The applications demonstrated in this article focused only on one facet of ||-coords, which is its graphical simplicity. We have only hinted at the possibilities opened up by the powerful geometrical properties of this representation: in particular, the fact that envelopes in the ||-coords representation enable direct identification of complex relationships such as linearity, coplanarity, and convex surfaces.

Another powerful property of ||-coords is the possibility of combining geometrical and nongeometrical axes in the same plot. In the context of conformational analysis this mixing has allowed us to apply double filtering and focus on conformations that share a common geometrical property and fall within a desired energy range. This combination of axes is, of course, not limited to angles and energy alone, but can be extended to include any relevant molecular property.

Acknowledgment

I thank Alfred Inselberg of many interesting discussions.

References

1. J. C. Hempel, R. M. Fine, M. Hassan, W. Ghoul, A. Guaragna, S. C. Koerber, Z. Li, and A. T. Hagler, *Biopolymers*, **36**, 282 (1995).
2. A. E. Howard and P. A. Kollman, *J. Med. Chem.*, **31**, 1669 (1988).
3. A. R. Leach, In *Reviews in Computational Chemistry*, Vol. 2, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, 1991.
4. I. D. Kuntz, E. C. Meng, and B. K. Shoichet, *Acc. Chem. Res.*, **27**, 117 (1994).
5. A. Inselberg, T. Chomut, and M. Reif, *J. Assoc. Comput. Mach.*, **34**, 765 (1987).
6. A. Inselberg and B. Dimsdale, *SIAM J. Appl. Math.*, **54**, 559 (1994).
7. A. Inselberg and B. Dimsdale, *SIAM J. Appl. Math*, **54**, 578 (1994).
8. J. C. Gower, *Biometrika*, **53**, 325 (1966).
9. O. M. Becker, *Proteins*, **27**, 213 (1997).
10. E. W. Basset, *Ind. Comput.*, **14**, 23 (1995).
11. B. T. Luke, *J. Chem. Inf. Comput. Sci.*, **33**, 135 (1993).
12. C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland, New York, 1991.
13. C. L. Brooks III, M. Karplus, and B. M. Pettit, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Vol. 71, John Wiley & Sons, New York, 1988.
14. F. H. Stillinger and T. A. Weber, *Phys. Rev. A*, **28**, 2408 (1983).
15. R. E. Bruccoleri and M. Karplus, *Biopolymers*, **29**, 1847 (1990).
16. Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84**, 6611 (1987).
17. H. Meirovitch and E. Meirovitch, *J. Comput. Chem.*, **18**, 240 (1997).
18. O. M. Becker and M. Karplus, *J. Chem. Phys.*, **106**, 1495 (1997).
19. R. E. Kunz and R. S. Berry, *J. Chem. Phys.*, **103**, 1904 (1995); K. D. Ball et al., *Science*, **271**, 963 (1996).
20. O. M. Becker, *J. Mol. Struct.* (*THEOCHEM*), **398–399**, 507 (1997).

21. L. S. D. Caves, J. Evenseck, and M. Karplus, in press.
22. R. Abagyan and P. Argos, *J. Mol. Biol.*, **225**, 519 (1992).
23. J. M. Troyer and F. E. Cohen, *Proteins*, **23**, 97 (1995).
24. R. Czerminski and R. Elber, *J. Chem. Phys.*, **92**, 5580 (1990).
25. E. G. Hutchinson and J. M. Thornton, *Prot. Sci.*, **5**, 212 (1996).
26. D. van der Spoel, B. L. De Groot, S. Hayward, H. J. C. Berendsen, and H. J. Vogel, *Prot. Sci.*, **5**, 2044 (1996).
27. W. Kabsch and C. Sander, *Biopolymers*, **22**, 2577 (1983).
28. P. E. Smith, H. D. Blatt, and B. M. Pettit, *Proteins*, **27**, 277 (1997).
29. G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.*, **7**, 95 (1963).
30. J. M. Thornton, In *Protein Folding*, T. E. Creighton, Eds., W. H. Freeman, New York, 1992, p. 59.
31. C. Wilson, M. R. Wardell, K. H. Weisgraber, R. W. Mahley, and D. A. Agard, *Science*, **252**, 1817 (1991).
32. M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. H. Kent, and A. Wlodawer, *Science*, **246**, 1149 (1989).
33. M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, *Acta Crystallogr. B*, **39**, 480 (1983).